A full-length bifunctional protein involved in c-di-GMP turnover is required for long-term survival under nutrient starvation in *Mycobacterium smegmatis*

Binod K. Bharati, Indra Mani Sharma, Sanjay Kasetty, Manish Kumar, Raju Mukherjee and Dipankar Chatterji

1Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India
2Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611-0700, USA

The bacterial second messenger cyclic diguanosine monophosphate (c-di-GMP) plays an important role in a variety of cellular functions, including biofilm formation, alterations in the cell surface, host colonization and regulation of bacterial flagellar motility, which enable bacteria to survive changing environmental conditions. The cellular level of c-di-GMP is regulated by a balance between opposing activities of diguanylate cyclases (DGCs) and cognate phosphodiesterases (PDE-As). Here, we report the presence and importance of a protein, MSDGC-1 (an orthologue of Rv1354c in *Mycobacterium tuberculosis*), involved in c-di-GMP turnover in *Mycobacterium smegmatis*. MSDGC-1 is a multidomain protein, having GAF, GGDEF and EAL domains arranged in tandem, and exhibits both c-di-GMP synthesis and degradation activities. Most other proteins containing GGDEF and EAL domains have been demonstrated to have either DGC or PDE-A activity. Unlike other bacteria, which harbour several copies of the protein involved in c-di-GMP turnover, *M. smegmatis* has a single genomic copy, deletion of which severely affects long-term survival under conditions of nutrient starvation. Overexpression of MSDGC-1 alters the colony morphology and growth profile of *M. smegmatis*. In order to gain insights into the regulation of the c-di-GMP level, we cloned individual domains and tested their activities. We observed a loss of activity in the separated domains, indicating the importance of full-length MSDGC-1 for controlling bifunctionality.

INTRODUCTION

Recent advances in the study of the social behaviour of bacteria motivated us to investigate quorum sensing and its role in bacterial survival. Quorum sensing facilitates the coordinated response of a bacterial population in response to environmental stress. In both Gram-negative and Gram-positive bacteria, small signalling molecules or autoinducers are the key components, whose presence can be sensed by neighbouring cells and which form the basis of cell–cell communication (Taga & Bassler, 2003; Reading & Sperandio, 2006). Bacterial biofilm formation is one of the best-known quorum-sensing phenomena (Bassler, 1999; Miller & Bassler, 2001). Although free-living bacteria can form biofilms in their natural habitats, biofilms can also be destabilized, and small molecules act as primary and secondary messengers, which are produced by bacteria to counter destabilization. Cyclic AMP (cAMP) and ppGpp are two well-known examples of second messengers which play important roles in relaying extracellular information (Chatterji & Ojha, 2001; Camilli & Bassler, 2006; Shenoy & Visweswariah, 2006). A novel second messenger, c-di-GMP [bis-(3’-5’)-cyclic diguanosine monophosphate], was reported in *Gluconacetobacter xylinus*, and acts as a positive allosteric regulator of the cellulose synthase pathway (Ross et al., 1987; Weinhouse et al., 1997; Tal et al., 1998). An increased level of c-di-GMP was found to stimulate the matrix (exopolysaccharide/cellulose) production required for biofilm formation, and was also found to regulate the transition between motile and sessile forms in Gram-negative bacteria (Kirillina et al., 2004; Hickman et al., 2005; Jenal & Malone, 2006; Kuchma et al., 2007). Various studies have established a clear link between the level of c-di-GMP and alterations in cell-surface properties, cell
differentiation, motility and biofilm formation in bacteria (D’Argenio et al., 2002; Tischler & Camilli, 2004; Kazmierczak et al., 2006; Beyhan et al., 2006; Tamayo et al., 2007; Merritt et al., 2007).

c-di-GMP is a condensation product of two GTP molecules, and it is hydrolysed to GMP via a linear intermediate product, pGpG. The synthesis and hydrolysis of c-di-GMP are carried out by diguanylate cyclase (DGC) and its cognate phosphodiesterase (PDE-A), respectively. The intracellular concentration of c-di-GMP is maintained by a balance between DGC and PDE-A activities. The GGDEF (Gly-Gly-Asp-Glu-Phe) and EAL (Glu-Ala-Leu) conserved motifs form major components of the two active sites in DGC and PDE-A, respectively. Mutations in these conserved residues result in reduced enzymic activities of the respective proteins (Paul et al., 2004; Christen et al., 2005). Studies have also shown the involvement of another domain, HD-GYP, with metal-dependent phosphodiesterase activity, which hydrolyses c-di-GMP to GMP (Galperin et al., 1999, 2001; Dow et al., 2006).

Studies demonstrating the role of c-di-GMP in various aspects of physiology have been mostly carried out in Gram-negative bacteria. Staphylococcus aureus is the only Gram-positive bacteria studied for c-di-GMP signalling in which c-di-GMP was found to inhibit cell–cell interactions and biofilm formation in a dose-dependent manner (Karaolis et al., 2005; Brouillette et al., 2005). Little is known, however, of c-di-GMP signalling in mycobacteria (Gupta et al., 2010). To characterize the physiological importance of c-di-GMP in mycobacteria, we undertook studies on MSMEG_2196 of Mycobacterium smegmatis, which encodes protein MSDGC-1. M. smegmatis is a non-pathogenic fast-growing species that is widely accepted as an excellent model system for studying various aspects of M. tuberculosis biology (Jacobs et al., 1987; Smuelders et al., 1999; Ojha et al., 2000; Dahl et al., 2005). MSDGC-1 has a tandem arrangement of GAF, GGDEF and EAL domains. We have systematically cloned, overexpressed and deleted msdgc-1 to understand its role in M. smegmatis. We have also shown that MSDGC-1 is a bifunctional protein with the ability to synthesize and hydrolyse c-di-GMP. Deletion of the msdgc-1 gene severely affects the long-term survival of mycobacteria under conditions of nutrient starvation (0.02 % glucose), and the full-length protein is required for DGC and PDE-A activities.

**METHODS**

**Bacterial strains, plasmids and growth conditions.** Plasmids and oligonucleotides used in this study are listed in Table 1 and Table S1 available with the online version of this paper. Escherichia coli strains DH5α and BL21(DE3) were grown in Luria–Bertani (LB) broth at 37 °C with agitation or on LB medium containing 1.5 % (w/v) agar. M. smegmatis mc²155 (Snapper et al., 1990) was grown in Middlebrook 7H9 (Difco) broth with 2 % (w/v) glucose as the carbon source and 0.05% (v/v) Tween 80, or on 7H9 medium containing 1.5 % (w/v) agar. Antibiotics were used at the following concentrations (μg ml⁻¹): ampicillin (100), kanamycin (35) or gentamicin (20) for *E. coli*, and kanamycin (20), gentamicin (10) or hygromycin (30) for *M. smegmatis*. PCRs were carried out using DyNAzyme EXT DNA Polymerase (Finzyme) following the manufacturer’s instructions. All clones generated were confirmed by sequencing (Microsynth, Switzerland, or MWG, India). Restriction enzymes used for cloning were obtained from New England Biolabs.

**Bioinformatics analysis.** The *M. smegmatis* genome was searched using the TIGR database (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi) for the annotation of proteins containing GGDEF and EAL domains.

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M. smegmatis</em> mc²155</td>
<td>Wild-type strain</td>
<td>Snapper et al. (1990)</td>
</tr>
<tr>
<td>DGCCKO</td>
<td><em>M. smegmatis</em> in which MSMEG_2196 has been replaced with the kan cassette, kan’</td>
<td>This study</td>
</tr>
<tr>
<td><em>M. smegmatis</em> pMV</td>
<td><em>M. smegmatis</em> transformed with pMV261 vector alone, kan’</td>
<td>This study</td>
</tr>
<tr>
<td>DGCCE</td>
<td><em>M. smegmatis</em> transformed with pMVDGC_2196, kan’</td>
<td>This study</td>
</tr>
<tr>
<td>DGCCO</td>
<td>DGCKO complemented with MSMEG_2196, kan’ hyg’</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC4K</td>
<td>Source of aph gene</td>
<td>Pharmacia Biotech</td>
</tr>
<tr>
<td>pET21b</td>
<td>Cloning vector (C-terminal His₆-tag), amp’</td>
<td>Novagen</td>
</tr>
<tr>
<td>pET28a</td>
<td>Cloning vector (N-terminal His₆-tag), kan’</td>
<td>Novagen</td>
</tr>
<tr>
<td>pPR27</td>
<td>Suicide vector, ger’</td>
<td>Pellic et al. (1997)</td>
</tr>
<tr>
<td>pMVDGC_2196</td>
<td>Cloning vector with hsp60 promoter, kan’</td>
<td>Stover et al. (1991)</td>
</tr>
<tr>
<td>pMVDGC_2196</td>
<td>Cloning vector with hsp60 promoter, hyg’</td>
<td>Stover et al. (1991)</td>
</tr>
<tr>
<td>pET DGC_2196</td>
<td>MSDGC-1 cloned in pMVDGC_2196, kan’</td>
<td>This study</td>
</tr>
<tr>
<td>pET DGC_2774 Nter</td>
<td>MSDGC-2 cloned in pET28a, kan’</td>
<td>This study</td>
</tr>
<tr>
<td>pGAF</td>
<td>GAF domain of MSDGC-1 cloned in pET21b, amp’</td>
<td>This study</td>
</tr>
<tr>
<td>pGGDEF</td>
<td>GGDEF domain of MSDGC-1 cloned in pET21b, amp’</td>
<td>This study</td>
</tr>
<tr>
<td>pEAL Nter</td>
<td>EAL domain of MSDGC-1 cloned in pET28a, kan’</td>
<td>This study</td>
</tr>
<tr>
<td>pGGDEF-EAL</td>
<td>GGDEF–EAL domain of MSDGC-1 cloned in tandem</td>
<td>This study</td>
</tr>
</tbody>
</table>
Two ORFs, MSMEG_2196 and MSMEG_2774, were annotated as putative GGDEF and EAL domain proteins. The protein sequence was subjected to Smart database (v.5.1; http://smart.embl-heidelberg.de/) and NCBI Conserved Domains database (v2.20; www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) searches to identify the various domains present in the protein. Furthermore, we did a BLAST analysis of the GGDEF and EAL domains in a non-redundant NCBI database to find protein homologues in other mycobacteria.

Cloning, expression and purification of MSDGC-1 and MSDGC-2. Genes MSMEG_2196 (1848 bp, 615 aa, MSDGC-1) and MSMEG_2774 (1080 bp, 359 aa, MSDGC-2) were PCR-amplified from M. smegmatis genomic DNA using primers MKE/MKR and MKDGC F/ MKDGC R, respectively (Table S1). The amplicons MSMEG_2196 and MSMEG_2774 were cloned into plastid pET21b using NheI/NcoI and Ndel/HindIII restriction sites, respectively. The pET28a vector was digested with the same set of restriction enzymes and used for N-terminal His-tag cloning of MSMEG_2774. The resulting plasmids pET DGC_2196, pET DGC_2774 and pET DGC_2774 Nter were transformed into E. coli BL21(DE3) for protein expression. Protein purification was done as described previously (Jain et al., 2006). Briefly, cells were induced with 1 mM IPTG for 3 h, harvested, and purified as described previously (Jain et al., 2006).

Enzymic assays. DGC and PDE-A assays were adapted from procedures described by others (Christen et al., 2005). Briefly, the purified proteins MSDGC-1, MSDGC-2 and PleD (5 μM) were incubated with 50 mM Tris/HCl (pH 7.9), 500 mM NaCl, 1 mM PMSF and 1 mg lysozyme ml^-1. After centrifugation, the lysozate was loaded onto a Ni-NTA column and washed with 100 column volumes of wash buffer containing 20 mM imidazole. The protein was eluted with elution buffer containing 500 mM imidazole, dialysed against a buffer containing 50 mM Tris/HCl (pH 7.9), 500 mM NaCl and 5 mM β-mercaptoethanol, and stored at 4 °C for future use.

Preparation of [α-32P]c-di-GMP and cold c-di-GMP. The protocol was adapted from that described by Christen et al. (2005) for the preparation of [α-32P]labelled and non-labelled c-di-GMP. For radiolabelled c-di-GMP, [α-32P]ATP (125 μCi (4.625 MBq); 3500 Ci mmol^-1 (129.5 TBq mmol^-1)) was incubated with 100 μg purified His6-tagged PleD protein containing 1 mM cold GTP. The protein was precipitated by boiling for 5 min followed by centrifugation and filtration through a 0.2 μm pore-size filter. The supernatant was tested for the presence of c-di-GMP by TLC. Non-labelled c-di-GMP was precipitated by boiling for 5 min followed by centrifugation at 12 000 r.p.m. (15 294 g) for 20 min. Supernatants (1 ml) were spotted onto PEI-cellulose plates and visualized.

Cloning, expression and purification of GAF, GGDEF and EAL domains of MSDGC-1. For the construction of GAF (amino acids 23–166), GGDEF (amino acids 175–334) and EAL (amino acids 348–590), pET DGC_2196 was used as a template and primers mkgAF (F)/mkgGDEF (R), mkkGDEF (F)/mkkGGDEF (R) and mkkEAL (F)/mkkEAL (R) (Table S1) were used, respectively. For the amplification of GGDEF–EAL, primers mkgGGDEF (F) and mkkEAL (R) were used. The required region was PCR-amplified and cloned into vector pET21b or pET28a at NheI and HindIII sites, resulting in pGAF, pGGDEF, pEAL and pGGDEF-EAL. The expression and purification of the various domains of MSDGC-1 were achieved using a strategy similar to that described above for the full-length protein, with some modifications, and activities were checked in vitro. For the in vivo activity assay, E. coli BL21 cells were transformed with full-length MSDGC-1 as well as its various domains and allowed to grow to OD600 0.2. Thereafter, cells were fed with [32P]Pi, HPO4^-2 and further allowed to grow till the OD600 reached 0.6, followed by induction with 1 mM IPTG. After further incubation for 3 h, cells were harvested and lysed with 11 M formic acid. Total nucleotide extracts were separated on a PEI-cellulose TLC plate and visualized.

Substrate binding and steady-state kinetics. To calculate the steady-state kinetic parameters Vmax and Km of the purified protein, MSDGC-1 was assayed for DGC activity, as the separate domains did not show any activity. The radiolabelled GTP was used as a standard. For the determination of the initial velocity, different concentrations of substrate (1–200 μM GTP) were used with same amount of protein (5 μM). The reaction was stopped after 30 min by boiling for 5 min followed by centrifugation at 12 000 r.p.m. at 4 °C for 20 min. Resulting supernatants (1 ml) were spotted onto PEI-cellulose plates, developed in 1 : 1.5 (v/v) (NH4)2SO4 and 1.5 M KH2PO4 (pH 3.6), and exposed as described above. The intensities of the spots were determined using the Multi Gauge v2.3 (Fujifilm) software from Science Lab and compared with the standard plot of radiolabelled GTP. Km and Vmax were determined from a linear plot of initial velocity (1/v), as a function of substrate concentration (1/[S]) and from a nonlinear Michaelis–Menten plot using GraphPad Prism (version 5.02).

Overexpression of MSDGC-1. The gene MSMEG_2196 was subcloned into the E. coli-Mycobacterium shuttle vector pMV261 (Stover et al., 1991) using primers Relco F and Relco R (Table S1), resulting in pMVDCG_2196. This vector was electroporated into M. smegmatis mc^155 and colonies were selected on 7H9 agar medium containing kanamycin. Growth of the overexpression strain (DGCOE) in liquid culture was followed by plating on an 7H9 agar plate containing 2 % (w/v) glucose as the carbon source.

Targeted replacement of MSMEG_2196 in M. smegmatis mc^155. In order to investigate the role of c-di-GMP in the physiology of M. smegmatis, gene MSMEG_2196 was deleted by allelic exchange. Conserved regions of GGDEF and EAL of MSDGC-1 were replaced with a kanamycin-resistance marker present in the suicide vector pPR27. A recombinant cassette was constructed to delete MSMEG_2196 from the chromosome of M. smegmatis mc^155 (Table 1). The cassette consisted of a 3.2 kb DNA fragment spanning from base 246 upstream of MSMEG_2196 to base 750 of MSMEG_2196 containing Xbal and EcoRI restriction sites, and a fragment spanning from base 1175 of MSMEG_2196 to base 321 downstream of MSMEG_2196 containing EcoRI and NotI restriction sites. The DNA fragment containing the kan gene was obtained by digesting the pUC4K vector with EcoRI and cloned between upstream and downstream regions of the gene using the same restriction site. After the preparative cloning steps, the whole recombinant cassette was transferred to the suicide delivery vector pPR27 to obtain the final construct pDGCKO. Competent M. smegmatis mc^155 cells were transformed with pDGCKO. The sucrose-resistant, gentamicin-sensitive and kanamycin-resistant colonies were selected at 39 °C for further analysis. Disruption of the gene and the recombination event were verified by Southern blot analysis and PCR of selected colonies. For PCR, primers MFK and MKR were used for confirmation using genomic DNA as a template. For Southern blot analysis, the genomic DNA of the putative mutant (DGCKO) as well as that of mc^155 was digested with PstI. The whole MSMEG_2196 gene (~1.8 kb from mc^155 was used as a probe and labelled with [α-32P]dATP by PCR. The complemented strain was generated by cloning a functional msdgc-1 gene into vector pYM361. Hereafter, all strains used will be described as follows: wild-type or M. smegmatis mc^155, the knockout strain as DGCKO, the overexpression strain as DGCOE and the complemented strain as DGCOO.
Extraction and detection of c-di-GMP from M. smegmatis. Wild-type, DGCKO and DGCCO cells were grown in 7H9 medium with 2.0 or 0.02 % (w/v) glucose at 37 °C with shaking. Cells were harvested at 24, 48 and 72 h and washed with sterile MilliQ water. Approximately 3 g of wet cell pellets was used to extract total nucleotides using a published method (Weinhouse et al., 1997). Briefly, cells were incubated on ice with 0.6 M perchloric acid (HClO4) and extracts were neutralized with 5 M potassium carbonate (K2CO3). Samples were centrifuged to remove the precipitate of KClO4, and the supernatant was lyophilized. Lyophilized extracts were resuspended in MilliQ water and subjected to HPLC separation. Purified c-di-GMP (obtained from Nanyang Technological University, Singapore) was used as a standard, and its retention time was used as the reference.

HPLC analysis and MS. The extracted samples (100 μl) were injected onto a C18 column (4.6 x 150 mm, Agilent Eclipse XDB-C18) and separated by reverse-phase HPLC (Agilent 1100). Buffer A (100 mM KH2PO4, 4 mM tetrabutyl ammonium hydrogen sulfate, pH 5.9) and buffer B (75 % (v/v) buffer A, 25 % (v/v) methanol) were used in the gradient program, as described by Ryjenkov et al. (2005). Nucleotides were detected at 254 nm. The HPLC fractions were desalted using a Sephadex G-10 column, lyophilized and spotted on a Nucleotides were detected at 254 nm. The HPLC fractions were used in the gradient program, as described by Ryjenkov et al. (100 mM KH2PO4, 4 mM tetrabutyl ammonium hydrogen sulfate, C18) and separated by reverse-phase HPLC (Agilent 1200). Buffer A was injected onto a C-18 column (4.6 x 150 mm, Agilent Eclipse XDB-C18) and separated by reverse-phase HPLC (Agilent 1100). Buffer A (100 mM KH2PO4, 4 mM tetrabutyl ammonium hydrogen sulfate, pH 5.9) and buffer B (75 % (v/v) buffer A, 25 % (v/v) methanol) were used in the gradient program, as described by Ryjenkov et al. (2005). Nucleotides were detected at 254 nm. The HPLC fractions were desalted using a Sephadex G-10 column, lyophilized and spotted on a MALDI-TOF MS plate using 4-chloro-4-nitrophenol (CNP) as matrix. MALDI-TOF analysis was performed in an Ultraflex TOF mass spectrometer (Bruker Daltonics) using the negative-ion detection mode.

Long-term starvation cultures. The wild-type, DGCKO and DGCCO strains were grown in 7H9 containing 0.02 % (w/v) glucose and 0.05 % (v/v) Tween 80. The antibiotics were omitted from the culture to rule out any effect that they might have on long-term survival. Bacterial cultures were declumped by vortexing the cells with 0.5 mm glass beads before plating on a 7H9 agar plate, as described elsewhere (Primm et al., 2000). The number of c.f.u. was determined at regular time intervals up to 20 days.

Biofilm formation and quantification assay. The biofilm formation assay was performed as described previously (Mathew et al., 2006; Naresh et al., 2010). Experiments were started from a culture of bacteria grown in 7H9 medium supplemented with 2 % glucose and 0.05 % Tween 80. The well-grown primary cultures (~OD600 1.5) of the wild-type, DGCOE, DGCKO and DGCCO were adjusted to OD600 0.5, and 10 μl volumes were inoculated into the centre of a plate. The spreading ability of mycobacteria was evaluated after incubation for 5–7 days at 37 °C in a humidified incubator.

RESULTS

Identification of GGDEF and EAL domain-containing proteins in M. smegmatis

Bioinformatics analysis revealed the existence of two genes, MSMEG_2196 (MSDG-1) and MSMEG_2774 (MSDG-2), encoding GGDEF–EAL and GGDEF domain proteins, respectively, when searches were performed against the M. smegmatis genome in the TIGR database. The domain organization of the sequences illustrated in Fig. 1(a) was obtained from the SMART database (http://smart.embl-heidelberg.de/smart/set_mode.cgi). The MSDGC-1 protein consists of GAF–GGDEF–EAL domains in tandem, while GGDEF is the only domain present in MSDGC-2. However, the N terminus of MSDGC-2 showed the presence of a transmembrane domain in the SMART database. Multiple

Fig. 1. (a) Representative domain architectures of proteins containing GGDEF, EAL and GAF motifs in M. smegmatis. The protein with GAF–GGDEF–EAL (MSDG-1) is present in almost all mycobacterial species. The domain symbols were taken from the SMART database. The ORF and number of amino acids in the sequence are given after the organism name. Predicted transmembrane regions are shown as rectangles. (b) Multiple sequence alignment showing the presence of the conserved GGDEF and EAL motifs in different species of mycobacteria.
Table 2. Proteins with GGDEF and EAL domains in different mycobacterial species.

- No proteins in this category.

<table>
<thead>
<tr>
<th>Organism (total no. of proteins with GGDEF and/or EAL domains)</th>
<th>No. of proteins with the domain organization:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GAF–GGDEF–EAL</td>
</tr>
<tr>
<td>M. smegmatis mc²155 (2)</td>
<td>1</td>
</tr>
<tr>
<td>M. bovis BCG Pasteur (1)</td>
<td>–</td>
</tr>
<tr>
<td>M. tuberculosis H37Rv (2)</td>
<td>1</td>
</tr>
<tr>
<td>M. tuberculosis CDC1551 (1)</td>
<td>1</td>
</tr>
<tr>
<td>M. leprae TN (4)</td>
<td>1</td>
</tr>
<tr>
<td>Mycobacterium sp. JLS (5)</td>
<td>1</td>
</tr>
<tr>
<td>Mycobacterium sp. MCS (9)</td>
<td>2</td>
</tr>
<tr>
<td>M. avium 104 (7)</td>
<td>2</td>
</tr>
<tr>
<td>M. avium subsp. paratuberculosis K-10 (6)</td>
<td>1</td>
</tr>
<tr>
<td>M. vanbaalenii PYR-1 (15)</td>
<td>1</td>
</tr>
<tr>
<td>M. gilvum PYR-GCK (22)</td>
<td>1</td>
</tr>
</tbody>
</table>

*GGDEF domain with one or more input sensory domain such as PAS, Rec, FlhA or GAF.
†GGDEF domain with conserved GEEF or degenerate SDSEF motif.
‡Unusual domain organization arrangement, with EAL at the N terminus of GGDEF or an AraH output domain in the case of M. leprae.
sequence alignment of GGDEF and EAL domains showed the presence of conserved signature motifs in various mycobacterial proteins, as depicted in Fig. 1(b). Table 2 shows the number and domain organization of different GGDEF and EAL domain-containing proteins in mycobacteria. Interestingly, most mycobacterial species appear to have at least one protein with the GAF–GGDEF–EAL domain organization, except for the *Mycobacterium bovis* BCG Pasteur strain. Few mycobacterial species appear to have a HD-GYP domain for c-di-GMP degradation (Table 2). Proteins from different mycobacterial species with GAF–GGDEF–EAL domain architecture share >60% amino acid identity among themselves.

*M. smegmatis* MSDGC-1 is a bifunctional protein in vitro

In order to investigate the regulation of c-di-GMP *in vitro*, MSDGC-1 and MSDGC-2 proteins were expressed in *E. coli* BL21(DE3) cells, and purified through an affinity column (Fig. 2) and checked for DGC and PDE-A activities. The MSDGC-1 protein showed both c-di-GMP synthesis and hydrolysis activities *in vitro* (Fig. 3a). The PEI-cellulose TLC was developed in 1:1.5 (v/v) saturated (NH₄)₂SO₄ and 1.5 M KH₂PO₄, pH 3.6, and a spot with the same Rᵣ value (0.2) as that of c-di-GMP was observed that was apparently followed by another spot with the Rᵣ value (0.4) of linear di-GMP (pGpG), as reported by Christen et al. (2006). A reaction mixture containing PleD protein, which is a known DGC protein of *Caulobacter crescentus*, was used as a positive control. PleD protein only shows c-di-GMP synthesis (Paul et al., 2004); however, c-di-GMP was followed by an additional spot (Fig. 3a) that was not detected among the reaction products with MSDGC-1.

For further confirmation of the dual activity, MSDGC-1 protein was incubated with cold GTP, and the reaction product was analysed for the presence of c-di-GMP and pGpG using negative-ion mode MALDI-TOF MS (Fig. S1) or subjected to HPLC (Fig. 3b), followed by collection of peaks and subsequent MALDI-TOF MS analysis. The peak at 16.8 min (Fig. 3b, 1) was that of standard GTP, with an *m/z* of 522 [M-H]⁻. Similarly, the peak at 19.2 min (Fig. 3b, 2) indicated the retention time of standard c-di-GMP and had an *m/z* value of 689.7 [M-H]⁻. The MSDGC-1 reaction product (30 min) was subjected to HPLC, and a peak at 19.07 min (Fig. 3b, 3), indicated the retention time of c-di-GMP and an *m/z* value of 689.6 [M-H]⁻. For the detection of pGpG (the hydrolysis product of c-di-GMP), the reaction mixture of MSDGC-1 was incubated for 45 min instead of 30 min, and we observed a peak at 16.8 min with an *m/z* value of 707.6 [M-H]⁻, corresponding to pGpG (Fig. 3b, 4). Thus, it is evident that MSDGC-1, like other orthodox DGCs and cognate PDE-A, can use GTP to synthesize c-di-GMP, which can subsequently be hydrolysed into pGpG. However, longer incubation of c-di-GMP with MSDGC-1 did not yield GMP. Partially purified fractions of N-terminal His₆-tagged MSDGC-2 did not show c-di-GMP synthesis *in vitro* (data not shown).

The full-length protein is required for DGC and PDE-A activity

The MSDGC-1 protein is a three-domain protein with GGDEF and EAL domains in tandem with an N-terminal GAF domain, and it possesses both DGC and PDE-A activities. The GAF (cGMP, adenylyl cyclase, FhlA) domain is a cyclic nucleotide-binding domain present in PDE-As (Hurley, 2003). To understand the regulation of MSDGC-1, we attempted to express GAF, GGDEF and EAL individually, and functionally characterized each one. However, we were not able to overexpress and purify the GAF domain, despite several attempts. The combined GGDEF–EAL, EAL and GGDEF domains were purified by the addition of Triton X-100 to the lysis buffer to improve solubility (Fig. 4a). We performed the DGC assay of the GGDEF protein with a normalized protein concentration, but did not observe a spot corresponding to c-di-GMP on the TLC plate, which indicates that the DGC activity of the GGDEF domain was absent (Fig. 4b, lane 3). Similarly, purified EAL domain protein did not show PDE-A activity, as there was no spot corresponding to pGpG (Fig. 4b, lane 4). This indicates that the full-length protein is required for both DGC and PDE-A activities. In addition, the combined GGDEF–EAL domain without GAF showed poor DGC activity and PDE-A activity was absent *in vitro* (Fig. 4b, lane 5). This suggests that the GAF domain is important for the activities of the combined domain. In order to avoid any *in vitro* experimental error, the activity assay was also performed in a heterologous system. We noticed that only
full-length MSDGC-1 protein was active in vivo (Fig. S2, lane 5). At this stage it remains unclear why the GGDEF–EAL domain protein did not show activity in this experiment in vivo. Nevertheless, from these experiments it appears that full-length MSDGC-1 is required for dual activity.

**Velocity (V\text{max}) and Michaelis constants (K\text{m}) of protein MSDGC-1**

MSDGC-1 is a bifunctional protein with the ability to synthesize and hydrolyse c-di-GMP. Since separate GGDEF and EAL domains were found to be inactive, quantification of their individual activities remained difficult. We quantified the K\text{m} and V\text{max} values for the DGC activity of protein MSDGC-1 in steady-state kinetic experiments using densitometric analysis. Fig. 5(a) shows the synthesis of c-di-GMP and its degradation as a function of increasing GTP concentration. The steady-state kinetic parameters were derived from Michaelis–Menten and Lineweaver–Burk plots, as shown in Fig. 5(b). The K\text{m} and V\text{max} values obtained from the Michaelis–Menten method were 22.3 μM and 0.99 ± 0.2 μM min\textsuperscript{-1}, respectively. Similarly, the K\text{m} and V\text{max} values obtained from the Lineweaver–Burk method were 25 μM and 0.95 ± 0.2 μM min\textsuperscript{-1}, respectively. The substrate (GTP) concentrations used in the Michaelis–Menten and Lineweaver–Burk methods were 1–200 and 5–200 μM, respectively. In this experiment, c-di-GMP was synthesized as a function of GTP concentration, and pGpG was produced as a function of increasing c-di-GMP concentration. K\text{m} and V\text{max} values obtained for c-di-GMP synthesis in our study followed Michaelis–Menten enzyme kinetics, and the cumulative rate due to c-di-GMP synthesis and degradation was used to calculate the initial velocity of c-di-GMP synthesis by MSDGC-1. Radiolabelled GTP was used to plot the standard curve. K\text{m} and V\text{max} values obtained in our study appear to be much higher than those reported by others (Christen et al., 2005). One of the reasons for the higher values obtained in our case could be the lack of modulators such as small molecules or activation through phosphorylation, which perhaps are necessary for the optimal activity of MSDGC-1.
Generation and characterization of a ΔMSMEG_2196 strain

In order to investigate the role of c-di-GMP in the physiology of M. smegmatis, the MSMEG_2196 gene was deleted by allelic exchange. Conserved GGDEF and EAL regions of MSDGC-1 were replaced with a kanamycin-resistance marker present in the suicide vector pPR27. Colonies were further tested by PCR and Southern blot analysis for confirmation of the crossover event. For PCR, primers MKF and MKR were used, and a shift of ~800 bp was observed in the case of the knockout strain (DGCKO) as compared with the wild-type (data not shown). During Southern blot analysis, the probe hybridized to a fragment of approximately 3.9 kb in the wild-type strain and one of approximately 4.7 kb in the knockout strain (DGCKO). The increase of ~800 bp in DGCKO resulted from the insertion of the kanamycin cassette (Fig. S3).

M. smegmatis has single protein that is known to contain a GAF–GGDEF–EAL domain; therefore, the resulting knockout strain should not contain any c-di-GMP (c-di-GMP). To test this feature, we analysed total nucleotide extracts by SDS-PAGE showing purified protein from an Ni-NTA column. Lanes: 1, protein marker; 2, GGDEF–EAL domain; 3, EAL domain; 4, GGDEF domain. The assay buffer used for determining DGC and PDE-A activity contained GTP and c-di-GMP as substrates, respectively, as described in Methods. Labelled c-di-GMP was prepared by incubating [α-32P]GTP with PleD, and was used without further purification.

Deletion of MSMEG_2196 affects long-term survival under nutrient starvation

The mutant (DGCKO) was assayed for its ability to survive for a prolonged period with a limited carbon source (0.02 % glucose). Bacterial cultures were incubated before plating to minimize errors in counting c.f.u. caused by clumping, and OD600 was also measured to monitor the growth of different variants in 0.02 % glucose. The knockout strain (DGCKO) had a reduced viability of more than 50 % after 5 days, with a persistent decrease in the number of viable cells as compared with the wild-type (Fig. 7a). To test whether inactivating MSDGC_2196 causes a polar effect, the DGCKO strain was complemented with the functional gene through the insertional plasmid pMV261 containing MSMEG_2196. The complemented strain (DGCCO) was able to survive carbon starvation similarly to the wild-type strain (Fig. 7a).

MSMEG_2196 was subcloned into a suicide vector, pMV261, under the control of the heat-shock promoter (Hsp60) for overexpression studies. The growth profile of the MSDGC-1-overexpressing strain (DGCOE) was compared with that of the wild-type in 7H9 broth supplemented with 0.02 % (w/v) glucose and 0.05 % (v/v) Tween 80. The overexpressing strain grew more slowly than the wild-type (Fig. 7b). On solid agar medium, colonies produced by overexpression and knockout strains showed different colony morphologies compared with the control (wild-type strain with pMV261 vector). The wild-type M. smegmatis showed colonies with a non-uniform margin and rugose surface (wrinkled). The overexpression strain produced a raised colony with smooth margins. The knockout strain had irregular-shaped colonies that were flat and had undulating margins. The complemented strain showed a colony morphology similar to that of the wild-type (Fig. 7c), although not identical. However, it should be mentioned here that phenotypic differences may also arise from the fact that the hsp60 promoter was used in this study, instead of the gene’s own promoter.
DISCUSSION

In this study, we detected the presence of an additional second messenger, c-di-GMP, and the protein responsible for its synthesis and hydrolysis in a Gram-positive bacterium, \textit{M. smegmatis}. Most of the known proteins with composite GGDEF and EAL domains show either DGC or PDE-A activity, whereby the inactive domain regulates the active partner. In the case of MSDGC-1, we separated and functionally characterized each of the individual domains \textit{in vitro}. We did not detect the formation of c-di-GMP or pGpG as reaction products when individual GGDEF and EAL domain proteins were used. We noticed a substantial reduction in DGC activity of the GGDEF–EAL domain in the absence of the GAF domain. Activities of separated domains were also checked in a heterologous system, which mimics \textit{in vivo} conditions. We were able to detect minimal c-di-GMP synthesis activity by full-length MSDGC-1 only. For the detection of c-di-GMP formation, an overexpression system was used, and any background c-di-GMP coming from the host with an empty vector remained undetectable (Fig. S2, lane 1) and acted as a control.

The requirement of a neighbouring domain or the full-length protein for complete activity has also been observed by others (Ryjenkov et al., 2005; Kazmierczak et al., 2006). MSDGC-1 consists of a GAF domain at its N terminus and is believed to modulate the activity of the protein with a GGDEF–EAL domain by binding to other molecules such as cAMP, cGMP and c-di-GMP. In proteins with a single GGDEF domain, the presence of a GAF domain or phosphorylation of the N-terminal receiver domain has been shown to increase the activity of the GGDEF domain (Paul et al., 2004; Ryjenkov et al., 2005). Removal of the GAF domain from MSDGC-1 renders the protein inactive, and this suggests that the domain has a role in the regulation of c-di-GMP synthesis and degradation. Recently, a more comprehensive understanding of the regulation of a bifunctional protein, Lpl0329, in the Gram-negative pathogenic bacterium \textit{Legionella pneumophila} has been obtained. In this case, phosphorylation-mediated control of dual activity has been demonstrated, in which upon phosphorylation, the DGC activity of the protein decreases but PDE-A activity remains (Levet-Paulo et al., 2011). While studying the bifunctional behaviour of MSDGC-1 \textit{in vitro}, we used the PleD reaction product to identify c-di-GMP as a positive control, since PleD is a bona fide DGC that synthesizes c-di-GMP. We consistently observed the presence of an additional spot in the PleD reaction product (indicated by an arrow in Fig. 3a) that was absent in the MSDGC-1 reaction product. This not only prompted us to identify the additional spot but also suggested that PleD and MSDGC-1 might use a different reaction mechanism to catalyse the synthesis of c-di-GMP from GTP as a substrate. In order to identify this spot, several experiments were carried out, and we concluded that the additional spot was a linear molecule.

\textbf{Fig. 5.} (a) Velocity ($V_{\text{max}}$) and Michaelis constants ($K_m$) of the MSDGC-1 protein. Synthesis of c-di-GMP as a function of increasing concentrations of GTP. Lane 1, control (assay buffer); lanes 2–18, increasing concentrations of GTP from 1 to 200 \textmu M. The concentration of the protein (MSDGC-1) was kept at 5 \textmu M. The PleD reaction product was used as a positive control (lane 19). (b) Michaelis–Menten (upper panel) and Lineweaver–Burk plots (lower panel) of c-di-GMP synthesis with increasing concentrations of GTP.
Fig. 6. Detection of c-di-GMP in *M. smegmatis* using reverse-phase HPLC (left panels) and MALDI-TOF (right panels). (a) Standard c-di-GMP showing the peak at 19.2 and m/z of 689.6; (b) total nucleotide extract from wild-type *M. smegmatis* grown for 72 h with 2% glucose (the presence of c-di-GMP is marked with an arrow); (c) absence of c-di-GMP in both reverse-phase HPLC and MALDI-TOF when the total nucleotide extract was analysed as in (b) but devoid of MSDGC-1; (d) reappearance of c-di-GMP in the complemented strain.
which is an intermediate product of c-di-GMP processing by the PleD protein (Fig. S5).

All mycobacterial species have an orthologue of MSDGC-1, which suggests that the protein may play similar roles in these species. Some species have more than one protein with a GGDEF or EAL domain, and the need for more than one protein for the modulation of c-di-GMP has been questioned by some. Studies done in different bacterial systems with multiple copies have suggested that this may help to maintain the level of c-di-GMP in localized compartments of the cell, providing bacteria with the ability to survive in different environmental niches (Römling et al., 2005). This appears to be true in case of the opportunistic pathogen Mycobacterium avium, which infects immunocompromised patients. The M. avium genome encodes a greater number of proteins with a GGDEF-EAL domain than those of M. smegmatis and M. tuberculosis. Other mycobacterial species, such as Mycobacterium gilvum and Mycobacterium vanbaalenii PYR-1, which have been isolated from sites of bioremediation of polyaromatic hydrocarbons, have even greater numbers of these proteins. This strongly suggests that c-di-GMP is involved in the survival and adaptation of mycobacteria. It has also been shown that during exponential growth, PDE-A activity is favoured to keep the c-di-GMP level low, and that during starvation, the level of c-di-GMP increases and as a result, c-di-GMP binds and activates the protein YcgR, which in turn binds to the flagellar motor directly and slows it down (Armitage & Berry, 2010; Merritt et al., 2010; Boehm et al., 2010; Paul et al., 2010).

In vivo detection of c-di-GMP in M. smegmatis was difficult using radiometric TLC techniques due to the low concentrations encountered. However, we successfully detected the presence of c-di-GMP in stationary phase cultures using HPLC coupled with MALDI-TOF MS. Interestingly, we observed a higher level of c-di-GMP (about sevenfold) in the case of the wild-type strain at stationary phase under carbon-starved conditions (0.02 % glucose) compared with a culture with 2 % glucose (Fig. S6). The level of c-di-GMP was normalized using an equal weight of wet cells, and purified c-di-GMP was used to generate a standard curve. Our data suggest that msdgc-1 is not involved in biofilm formation and that an increased level of c-di-GMP is required in the M. smegmatis stationary phase for survival under nutrient-depleted conditions. Perhaps it is not out of context to mention that the level of the stringent factor ppGpp is maintained by two opposing activities of a single polypeptide, Rel, in Gram-positive bacteria (Chatterji &

![Fig. 7. Role of msdgc-1 in long-term survival and phenotypic characterization. (a) Long-term survival analysis by c.f.u. determination; (b) growth profile; (c) colony morphology. Strains used in this study were: wild-type, msdgc-1 deleted strain (DGCKO), msdgc-1 complemented strain (DGCCO), and msdgc-1 overexpression strain (DGCOE), where required. (a) and (b) were performed under carbon-limited conditions (0.02 % glucose), and (c) was performed under carbon-replete conditions (2 %) glucose. All experiments were done in triplicate.](http://mic.sgmjournals.org)
ACKNOWLEDGEMENTS

We thank Professor Urs Jenal, Biozentrum, Switzerland, for providing the PleD clone for the c-di-GMP assay. This work was financially supported by DBT, Government of India. B.K.B. acknowledges IISc, Bangalore, for a fellowship. I.M.S. and R.M. acknowledge CSIR for fellowships. S.K. and M.K. acknowledge DBT for postdoctoral fellowships. The authors would also like to record their appreciation to Y. S. Prakash, Sathisha and Sunita for help in MS, and Dhanarajan T. for help in desalting the reverse-phase HPLC fractions.

REFERENCES


Edited by: T. Parish